基于对比增量学习的细粒度恶意流量分类方法

王一丰[1]，郭渊博[1]，陈庆礼[1]，方晨[1]，林韧昊[2]，周永良[1]，马佳利[1]
（1. 信息工程大学密码工程学院，河南 郑州 450001；2. 郑州大学计算机与人工智能学院，河南 郑州 450001）

摘 要：为应对层出不穷的新型网络威胁，提出了一种基于对比增量学习的细粒度恶意流量识别方法。所提方法基于变分自编码器和极值理论，在对已知类、小样本类和未知类流量实现高性能检测的同时，还可以在不采用大量原任务样本的条件下快速实现对新增恶意类的识别，以满足增量学习场景下对存储成本和训练时间的要求。具体来说，模型将对比学习融入变分自编码器的编码阶段，并采用 A-Softmax 实现对已知类和小样本类的识别；将变分自编码器重构与极值理论结合，采用重构误差实现对未知类的识别；利用变分自编码器存储原有类知识，采用样本重构和知识蒸馏方法，在不采用大量原有类样本的条件下实现对所有类样本的识别。实验结果表明，所提方法不仅实现了对已知类、小样本类和未知类流量高性能检测，并且所设计的样本重构和知识蒸馏模块均可有效降低增量学习场景下对原有类知识的遗忘速度。

关键词：网络流量分类；变分自编码器；增量学习；对比学习
中图分类号：TP393
文献标志码：A
DOI: 10.11959/j.issn.1000–436x.2023068

Method based on contrastive incremental learning for fine-grained malicious traffic classification

WANG Yifeng[1], GUO Yuanbo[1], CHEN Qingli[1], FANG Chen[1],
LIN Renhao[2], ZHOU Yongliang[1], MA Jiali[1]
1. Cryptography Engineering Institute, Information Engineering University, Zhengzhou 450001, China
2. College of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China

Abstract: In order to protect against continuously emerging unknown threats, a new method based on contrastive incremental learning for fine-grained malicious traffic classification was proposed. The proposed method was based on variational auto-encoder (VAE) and extreme value theory (EVT), and the high accuracy could be achieved in known, few-shot and unknown malicious classes and new classes were also identified without using a large number of old task samples, which met the demand of storage and time cost in incremental learning scenarios. Specifically, the contrastive learning was integrated into the encoder of VAE, and the A-Softmax was used for known and few-shot malicious traffic classification, EVT and the decoder of VAE were used for unknown malicious traffic recognition, all classes could be recognized without a lot of old samples when learning new tasks by using VAE reconstruction and knowledge distillation methods. Experimental results indicate that the proposed method is efficient in known, few-shot and unknown malicious classes, and has greatly reduced the forgetting speed of old knowledge in incremental learning scenarios.

Keywords: network traffic classification, variational auto-encoder, incremental learning, contrastive learning

收稿日期：2022–10–20；**修回日期**：2023–01–18
通信作者：郭渊博，yuanbo_g@hotmail.com
基金项目：国家自然科学基金资助项目（No.62276091）；河南省重大公益专项基金资助项目（No.201300311200）
Foundation Items: The National Natural Science Foundation of China (No.62276091), Major Public Welfare Project of Henan Province (No.201300311200)

0 引言

网络入侵检测系统（NIDS, network intrusion detection system）已经成为信息系统中检测网络攻击的重要手段。随着物联网（IoT, Internet of things）等技术的快速发展与应用，信息系统中的终端设备数量不断增加。这些终端设备往往资源受限，且由于固件更新缓慢、难以部署安全代理等原因，更易遭受攻击。这种趋势进一步加剧了基于流量的NIDS的重要性。

NIDS 的任务是对恶意流量进行识别和分类。现有流量分类方法主要有基于端口、基于载荷和基于流三类[1]。然而，前 2 种方法难以应对如今愈发智能的网络攻击。由于端口跳变等技术的存在，基于端口的方法的精度大幅降低[2]。另一方面，加密技术使流量载荷特征难以获取，难以应用基于载荷的方法[3]。而基于流的方法具有较强的适用性，特别是近几年机器学习和深度学习取得了跨越式发展，使这类方法性能得到了显著提高[4]。

现有基于机器学习的检测方法虽然表现优异，但其训练过程受制于固定数据。实际中各种新类型应用或威胁持续涌现，产生了各种新类型的良性或恶意流量。此外，在恶意流量识别中，更细粒度的分类结果能够为安全专家提供更详细的威胁相关信息，有助于更快实施响应措施。现有基于机器学习的流量分类方法面临如下挑战。

1) 流量分类是一个开集识别问题。愈发常见的未知网络攻击（如零日攻击、新恶意软件以及针对新兴技术的攻击等）产生了各类未知恶意流量，NIDS 必须具有检测未知恶意流量的能力[5]。现有研究多在闭集环境下进行，缺乏对未知恶意流量的检测。

2) 流量分类是一个增量学习问题，也称终生学习或持续学习问题[6]。不断新增的各类型流量需要 NIDS 持续细粒度识别。在这种需求下，检测模型需要满足：当新类别在任意时间出现时，模型都是可训练的。而传统机器学习方法若直接处理分布变化的数据，会出现灾难性遗忘问题[7]，即学习新类别时会迅速遗忘原有类知识。在实际应用中，以往方法（联合学习）需要维持包含所有原有类的庞大数据集或保留所有原任务的模型参数，导致存储开销和数据投毒等风险不断增加。并且每次学习识别新增恶意类流量时，都需要重新训练网络模型参数，导致时间、维护和存储开销越来越大，难以满足在大规模网络系统中对识别效率的要求。当前研究普遍缺乏对增量学习场景的考虑。

3) 流量分类是一个细粒度小样本（C2FS, coarse-to-fine few-shot）[8]问题。深度学习的分类性能很大程度上取决于训练数据的质量和数量[9]，然而实际中常常缺乏足够的标注训练数据，以往方法大多没有考虑这种需求。

针对上述问题，本文提出了一种基于对比学习[10]和知识蒸馏[11]的细粒度增量恶意流量识别方法，主要贡献如下。

1) 结合团队研究成果[12]，本文将对比学习、变分自编码器（VAE, variational auto-encoder）[13]、A-Softmax[14]以及极值理论（EVT, extreme value theory）[15]结合，使所提方法能同时实现对已知类、小样本类和未知类恶意流量的细粒度分类。

2) 采用知识蒸馏方法保留已学习过的原有类知识相比，采用 VAE 压缩存储原任务上的样本信息，使所提方法具有增量学习能力的同时，时间和存储开销随新任务缓慢线性增长。

3) 在公开数据集上的实验结果表明，与其他方法相比，所提方法不仅实现了对已知类、小样本类和未知类流量的细粒度分类，并且在连续学习新任务时对原有类知识的遗忘速度较低，实现了增量学习场景下的细粒度流量识别。

1 相关工作

1.1 开集恶意流量识别

随着系统数据规模的不断增加以及加密技术的滥用，机器学习，特别是深度学习在细粒度恶意流量识别中表现优异，具有使用场景广、分类准确率高、加密流量可识别等优点[16]。

然而，这类方法难以应对越来越常见的未知攻击，并且与其他领域的开集识别问题不同，恶意流量识别不仅需要识别未知类样本，还需判断未知类样本是新的恶意流量还是新的良性流量。目前，开集恶意流量识别相关研究较少，其中 EVT 是最常用的方法。文献[17]采用威布尔校准支持向量机模型实现对已知类流量的细粒度识别后，采用 EVT 进一步发现未知类流量。文献[18]采用 EVT 计算各已知类的边缘距离分布以实现开集恶意流量识别。文献[5]设计了两阶段方法，采用条件变分自

编码器（CVAE, conditional VAE）和 EVT 方法分别实现对已知类和未知类流量的识别。但目前在该领域中，对未知类恶意流量的识别精度还远远不能满足应用需求。

1.2　增量学习

增量学习旨在解决机器学习中的灾难性遗忘问题，即模型在新任务上训练时在原任务上的性能通常会显著下降。原因在于一般机器学习模型假设数据分布是平稳的，即训练时接受同分布的数据训练。但如果训练数据分布是非平稳的，模型在新任务上训练时则会倾向于新任务而遗忘原任务，从而导致模型在原任务上的性能下降。文献[7]对增量学习提出了 3 个需求：1) 新类别在任何时间出现时，模型都应是可训练的；2) 在任何时间，模型都应对已学习过的所有类有较好的识别效果；3) 模型的计算和存储开销应是有限的，并随着类别数量的增加缓慢增长。

增量学习方法主要基于微调，可以分为基于参数隔离的方法、基于正则化的方法和基于回放的方法[7]。其中，基于参数隔离的方法[19]类似于多任务学习，模型在学习新任务时保留原任务的部分参数训练，通过为不同任务分配独立的参数空间，将新任务和原任务的参数相互隔离，从而实现增量学习。显然，这类方法极大地限制了新任务的数量和顺序，因此现有研究更多关注其他 2 种方法。

基于正则化的方法[20-22]通过对新任务损失函数施加不同约束实现，如修正更新梯度以保护原有类知识不被新增类知识覆盖。但这类方法高度依赖于新任务与原任务之间的相关性。

基于回放的方法[23-24]通过保留部分有代表性的原有类样本或高级特征表示实现，在学习新任务时这些原有类样本信息会一同训练以帮助模型兼顾原任务。这类方法需要额外的计算和存储开销，当任务种类不断增多时，要么训练成本增加，要么原有类知识的代表性减弱，同时实际使用中还可能存在隐私问题。

综上，现有的增量学习方法很难直接应用于恶意流量检测，原因如下。

1) 需要应对相似恶意类的挑战。由于网络攻击的对抗性，恶意类倾向于将自身特征模仿良性类规避检测，使新增类识别困难。

2) 需要应对不断新增的未知类流量的挑战。很多新型的未知类恶意流量缺乏标记数据，并且良性类模式也可能随时间改变，导致对未知类识别困难。

3) 需要更快的检测时间。在实时吞吐量较大的大规模系统中，NIDS 需要能快速检测和响应，而复杂模型的检测精度虽然更高，但训练时间较长，难以满足增量学习条件下的检测需求。

1.3　对比学习

对比学习的目标是学习一种数据变换方式，使特征空间中同类样本接近，异类样本远离，从而使下游分类任务更容易被解决。对比学习方法大多基于负样本的对比损失实现，分类性能主要取决于负样本的数量和质量[25]。现有 3 种主流对比方法如下。

1) 以 SimCLR[26]为代表的方法。这类方法随机选取当前训练批次中的异类样本作为负样本，简单高效。

2) 以 MoCo[27]为代表的方法。这类方法通过维护一个先进先出大批负样本队列，每次训练只更新最旧的一小批负样本。模型采用模板网络实现特征提取，并基于动量缓慢更新参数。

3) 以 AdCo[28]为代表的方法。这类方法设计负样本网络表示整体负样本，以生成高质量的负样本。

据本文所知，除了团队之前的研究[12]，目前还没有其他将对比学习应用于流量分类的研究，这主要是因为对比学习主要作用于提升小样本类的识别能力，但会增加模型训练时间。

2　方法设计

现有基于深度学习的 NIDS 难以快速实现对新类样本的识别。为此，本文提出了一种基于对比增量学习的细粒度恶意流量分类方法，实现对已知类、未知类和新增类恶意流量的细粒度识别。

该方法采用 CICFlowMeter[29]等工具提取的流特征实现分类。流特征经预处理后记作 d 维实值特征向量 $\boldsymbol{x}_i \in \mathrm{R}^d$，类标记为 y_i。原任务已知类标记集合为 $C_{\text{old}} = \{B, M_1, \cdots, M_k, F_1, \cdots, F_f\}$，其中，$B$ 表示良性类，$M_1, \cdots, M_k$ 表示 k 个大样本已知类，$F_1, \cdots, F_f$ 表示 f 个小样本已知类。设 U 为未知类，当出现新任务时，定义新增已知类标记集合为 $C_{\text{new}} = \{M_{k+1}, \cdots, M_{k+\Delta k}, F_{f+1}, \cdots, F_{f+\Delta f}\}$，其中，$\Delta k$ 和 Δf 分别代表新增的已知类和小样本类的数量，则当前所有类标记集合记为 $C_{\text{all}} = C_{\text{new}} \cup C_{\text{old}} \cup \{U\}$。令 $\boldsymbol{x}_o \in \{\boldsymbol{x}_i \mid y_i \in C_{\text{old}}\}$ 表示原任务上的样本，

$\boldsymbol{x}_n \in \{\boldsymbol{x}_i \mid y_i \in C_{\text{new}}\}$ 表示新任务上的样本，则所提方法的总体目标是在不采用或只采用较少 C_{old} 类样本 $\boldsymbol{x}_o$ 的条件下，模型能迅速实现对所有已知类、未知类以及在任意时间出现的新增类的样本的细粒度分类。

所提方法的检测模型如图 1 所示。模型分为原任务和新任务上的恶意流量识别两部分，且这两部分模型结构相同，均包含已知类识别阶段和未知类识别阶段。首先，在原任务上，模型采用 C_{old} 类样本 $\boldsymbol{x}_o$ 及其对应标记进行训练。其次，在已知类恶意流量识别阶段，模型结合对比损失训练 VAE，使异类样本在潜特征空间中进一步区分，采用 A-Softmax 将潜特征 $\boldsymbol{z}$ 映射到分类空间中，用分类损失函数训练模型；在未知类恶意流量识别阶段，模型采用重构损失训练 VAE 解码器，并用 EVT 实现对未知类的识别。最后，在新任务阶段，模型先从原模型中继承初始参数，并采用知识蒸馏和原 VAE 模型存储的原有类知识，与新增类样本一同基于分类损失训练模型的网络参数，从而实现增量学习场景下的恶意流量识别。

2.1 已知类恶意流量识别

该阶段对应图 1 中的原任务上恶意流量识别中的 VAE 和 A-Softmax 部分，目标是采用 C_{old} 类样本及其对应标记训练模型，使其对所有 C_{old} 类样本的预测结果都尽量接近其真实标记。

为此，本文采用 VAE 模型架构实现。因为本文发现采用对比学习结合 A-Softmax 作为分类函数时，其分类精度相当且训练速度更快，所以不采用之前研究中分类效果更优秀的 CVAE 模型。并且，由于在增量学习需求下标记 y 的向量维度会改变，因此 VAE 更适应于此场景。

首先，模型先将样本 $\boldsymbol{x} \in \mathrm{R}^d$ 转化为更低维的潜特征 $\boldsymbol{z} \in \mathrm{R}^h, h \ll d$，在更低维的特征空间中计算速度更快且更能区分不同类样本[5]。VAE 目标是生成接近但不同于输入 $\boldsymbol{x}$ 的重构样本 $\boldsymbol{x}'$。首先，对于给定 $\boldsymbol{x}$，其潜特征 $\boldsymbol{z}$ 的后验分布记为 $p(\boldsymbol{z} \mid \boldsymbol{x})$，并假设其服从正态分布，采用 $q(\boldsymbol{z} \mid \boldsymbol{x})$ 近似模拟。VAE 模型的目标是最大化其证据下界（ELBO, evidence lower-bound）[30]，即

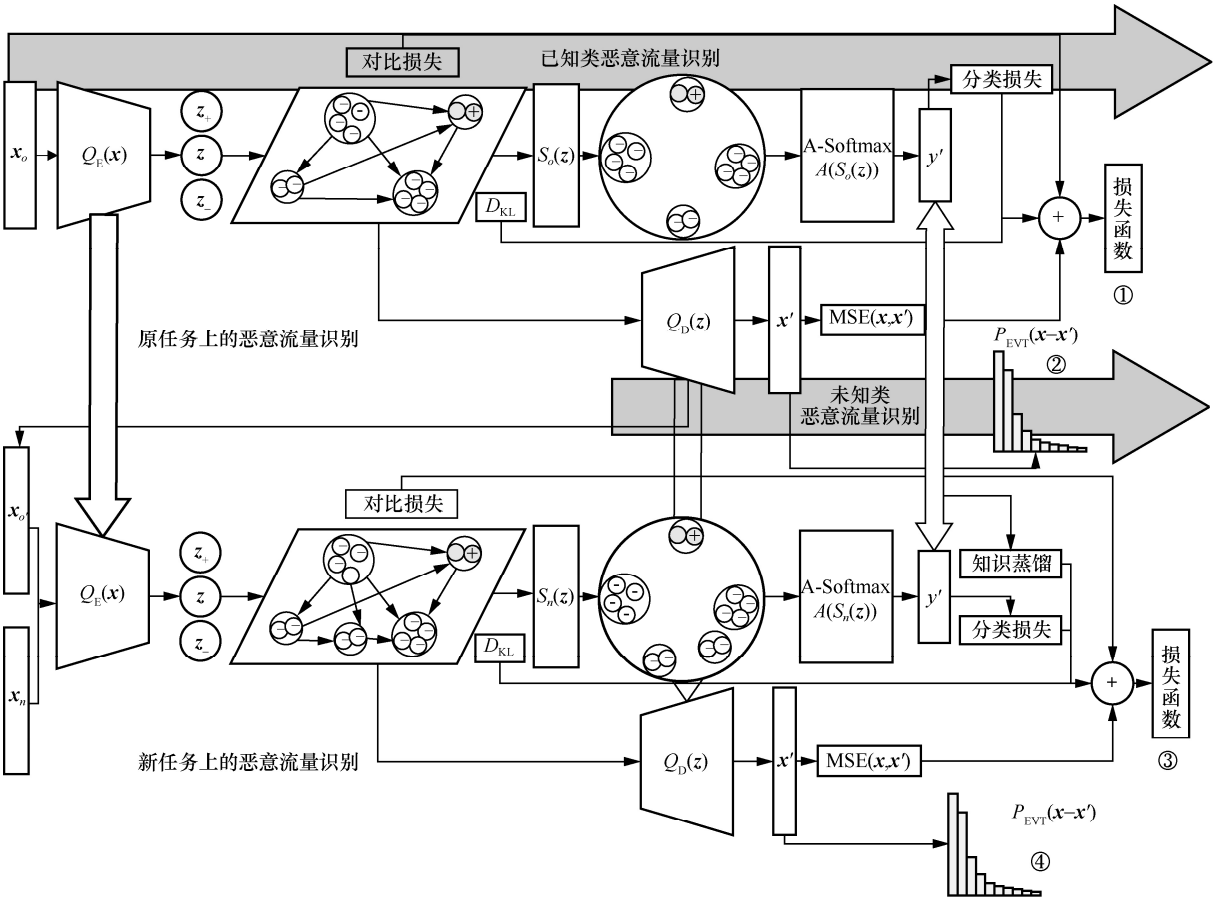


图 1 所提方法的检测模型

$$L_{\mathrm{ELBO}}=-\mathbb{E}_{q(z|\boldsymbol{x})}\left(\log(P(\boldsymbol{x}\mid\boldsymbol{z}))\right)+D_{\mathrm{KL}}\left(q(\boldsymbol{x}\mid\boldsymbol{z})\,\|\,p(\boldsymbol{z})\right)\tag{1}$$

其中，$P(\boldsymbol{z})$ 表示标准正态分布，D_{KL} 表示 KL（Kullback-Leibler）散度[31]。VAE 的损失函数可表示为

$$L_{\mathrm{vae}}=-\mathbb{E}_{q(z|x)}\left(\log(P(\boldsymbol{z}\mid\boldsymbol{x}))\right)+\mathrm{D}_{\mathrm{KL}}\left(N(\mu,\sigma)\,\|\,N(0,1)\right)\tag{2}$$

其中，μ和σ 分别为$q(\boldsymbol{z}\mid\boldsymbol{x})$ 的均值和标准差，计算时 VAE 采用编码网络$Q_{\mathrm{E}}(\boldsymbol{x})$ 近似$q(\boldsymbol{z}\mid\boldsymbol{x})$，解码网络$Q_{\mathrm{D}}(\boldsymbol{z})$ 近似模拟 $p(\boldsymbol{z}\mid\boldsymbol{x})$。实际损失函数的第一项采用原样本 $\boldsymbol{x}$ 与重构样本 $Q_{\mathrm{D}}\left(Q_{\mathrm{E}}(\boldsymbol{x})\right)$ 的均方误差（MSE, mean square error）衡量。

其次，为实现流量分类，本文在损失函数中引入对比学习，使特征空间中同类样本尽可能接近，异类样本尽可能远离。对比学习可以帮助提高分类准确率[27]。特别是 MoCo 类方法，由于其负样本数量很多，对小样本类提升很大[12]。对于任意样本 $\boldsymbol{x}_i$，$Q_{\mathrm{E}}(\boldsymbol{x}_i)$ 应与同类样本 $\boldsymbol{x}_+$ 的 $\boldsymbol{z}_+=Q_{\mathrm{E}}(\boldsymbol{x}_+)$ 接近，与异类样本 $\boldsymbol{x}_-$ 的 $\boldsymbol{z}_-=Q_{\mathrm{E}}(\boldsymbol{x}_-)$ 远离。对比学习的损失函数采用 InfoNCE 损失[32]的形式表示为

$$L_{\mathrm{Cont}}=-\mathbb{E}\left(\log\frac{\exp\left(Q_{\mathrm{E}}(\boldsymbol{x}_i)\dfrac{Q_{\mathrm{E}}(\boldsymbol{x}_+)}{\tau}\right)}{\sum\exp\left(Q_{\mathrm{E}}(\boldsymbol{x}_i)\dfrac{Q_{\mathrm{E}}(\boldsymbol{x}_-)}{\tau}\right)}\right)\tag{3}$$

其中，τ 是归一化超参数。

在原任务中，计算对比损失时负样本采用了不同的选取策略：对于大样本类$(B,M_1,\cdots,M_k)$，其负样本采用简单 SimCLR 策略随机选择；对于小样本类$(F_1,\cdots,F_f)$，采用 MoCo 策略选择尽可能多的负样本。

在分类器的选择上，传统神经网络一般采用 Softmax 实现。但由于模型中采用的对比损失几何意义上是基于余弦距离来衡量输入向量之间的相似度，因此在分类时采用角投影层函数 S_o 先将潜特征 z 投影到超球面上，并采用 A-Softmax 函数 A 基于角距离最终分类，则分类损失函数表示为

$$L_{\mathrm{Ang}}=-\mathbb{E}\left(\log\frac{\exp\left(\left\|Q_{\mathrm{E}}(\boldsymbol{x}_i)\right\|\psi(\theta_{S_o(y_i),S_o(Q_{\mathrm{E}}(\boldsymbol{x}_i))})\right)}{\exp\left(\left\|Q_{\mathrm{E}}(\boldsymbol{x}_i)\right\|\psi\left(\theta_{S_o(y_i),S_o(Q_{\mathrm{E}}(\boldsymbol{x}_i))}\right)\right)+\sum\limits_{j\neq y_i}\exp\left(\left\|Q_{\mathrm{E}}(\boldsymbol{x}_i)\right\|\cos(\theta_{S_o(y_i),S_o(Q_{\mathrm{E}}(\boldsymbol{x}_i))})\right)}\right)\tag{4}$$

其中，θ 表示下标的 2 个向量之间的夹角，并且对于任意的$\theta\in\left[\dfrac{k\pi}{m},\dfrac{(k+1)\pi}{m}\right]$，$\psi(\theta)=(-1)^k\cos(m\theta)-2k$，$m$ 为整数超参数（实验中取值为 4）。实验发现，对比学习的加入实际上更有利于角投影层函数 S_o 和 A-Softmax 函数 A 的参数训练，相反如果不采用对比学习，S_o 和 A 的训练难度在部分类上显著增加。

最后，训练时模型采用随机梯度下降算法更新网络参数，训练数据为原任务上的C_{old}类样本 x_o，其损失函数如式(5)所示，对应于图 1 中的①。

$$L_{\mathrm{All}}=L_{\mathrm{vae}}+\alpha L_{\mathrm{Ang}}+\beta L_{\mathrm{Cont}}\tag{5}$$

其中，α 和 β 是权重超参数。

2.2　未知类恶意流量识别

该阶段对应图 1 中的原任务上的恶意流量实别中的 EVT 部分，目标是对未知类（U）样本进行识别，本文采用 VAE 解码器结合 EVT 实现。

在 VAE 模型完成 2.1 节的训练后，不同已知类样本在编码后的特征空间中应当有明显区分。若样本 $\boldsymbol{x}_i$ 经过分类模型预测的标记 y_i' 正确且被分类为已知大样本类，即 $y_i'=y_i\in\{B,M_1,\cdots,M_k\}$，则在对比学习策略下其编码得到的潜特征$Q_{\mathrm{E}}(x_i)$ 应接近 y_i' 类的潜特征中心$\bar{z}_{y_i'}$，并且其重构样本x' 应接近 y_i' 类中心的重构样本$Q_D\left(\bar{z}_{y_i'}\right)$。基于此思路，本文为当前每个已知大样本类（$B,M_1,\cdots,M_k$）都构建了 k+1 个独立的 d 维 EVT 模型来实现对未知类的识别。

EVT 认为对于任意随机变量 X，其极值相对于阈值 t 的超出部分应服从广义帕累托分布（GPD, generalized Pareto distribution），即

$$P(X-t>x\mid X>t)\sim\left(1+\frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}\tag{6}$$

其中，$\gamma,\sigma>0$。实际计算时可以采用极大似然估计方法计算γ,σ，对数似然函数表示为

$$\log L(\gamma,\sigma)=-N_t\log\sigma-\left(1+\frac{1}{\gamma}\right)\sum_{i=1}^{N_t}\left(1+\frac{\gamma\left(X_i-t\right)}{\sigma}\right)\tag{7}$$

其中，N_t 表示观测数据中超过阈值 t 的样本数量，X_i 表示观察数据。本节采用已知大样本类样本与其类中心的重构误差$\left|Q_{\mathrm{D}}\left(Q_{\mathrm{E}}(x_i)\right)-Q_{\mathrm{D}}\left(\bar{z}_{y_i}\right)\right|$作为观察数据，计算当前 y_i 类的 EVT 模型参数，对应于图 1 中的②。

测试阶段模型的总体流程如图 2 所示。测试时，先将输入样本 x_j 由分类模型预测得到其分类标记 y_j'，若 $y_j' \in \{F_1,\cdots,F_f\}$ 为小样本类，分类结果为 y_j'；若 $y_j' \in \{B,M_1,\cdots,M_k\}$，则继续由 y_j' 类的 EVT 模型分析其重构样本与 y_j' 类中心重构样本的差异，即 $P_{\mathrm{EVT}}^{y_j'}\left(\left\|Q_{\mathrm{D}}(Q_{\mathrm{E}}(x_{\mathrm{j}}))-Q_{\mathrm{D}}\left(\bar{z}_{y_j'}\right)\right\|\right)$，判断其分类正确的概率。若其概率小于阈值，则最终识别结果为未知类 U；否则，分类结果仍为 y_j'。

2.3 新增类恶意流量识别

该阶段对应图1中新任务上的恶意流量检测部分，目标是在原任务上的恶意流量检测的基础上，不采用大量 C_{old} 类样本训练新的分类模型，使该新模型能对所有 C_{all} 类样本的预测结果都尽量接近其真实标签。

新模型的结构以及训练流程与原模型基本相同，不同之处在于标签向量 y 的维度（原任务上标签维度为 $\|C_{\mathrm{old}}\|$，新任务上标签维度为 $\|C_{\mathrm{old}} \cup C_{\mathrm{new}}\|$）。从图 1 中可以看出，该参数只影响角投影层函数 S。因此，新模型其他部分参数可以直接继承原模型的参数。而本文希望尽量保持原模型的知识，因此训练时单独对新的角投影层函数 S_n 采用了更大的学习率。

为了识别原任务上的 C_{old} 类样本，本文提出了基于 VAE 重构的模块和基于知识蒸馏的模块。首先，对于 C_{old} 中的小样本类，其本身样本数量较少因此不会占用很多存储和计算资源；对于良性类 B，其能随时轻易获取也不需要存储。在此场景下，小样本类和良性类样本可以直接应用于新任务。而本节的重点在于如何转移 C_{old} 类中大样本恶意类知识。

基于 VAE 重构的模块采用原模型 VAE 的解码器生成 C_{old} 类别的示例样本。由于 VAE 能够生成与原样本接近但不同的样本，因此可以利用原 VAE 模型实现原任务中的大样本恶意类样本 $\boldsymbol{x}_o$ 的重构。具体而言，在模型训练完成后对所有 $y_i \in \{M_1,\cdots,M_k\}$ 的样本计算编码后的潜特征 $\boldsymbol{z}$。对于这些潜特征 $\boldsymbol{z}$，其在对比学习作用下同类应互相接近，异类应互相远离。原模型对同属于 y_i 类样本的潜特征分别计算均值 $\bar{\boldsymbol{z}}_{y_i}$ 和标准差 $\sigma_{z_{y_i}}$，并采用高斯分布 $N(\bar{\boldsymbol{z}}_{y_i},\lambda\sigma_{z_{y_i}})$ 存储每个大样本恶意类的知识，其中，$\lambda \in [0.5,1]$ 为超参数，用于缩小标准差以防生成的样本过度偏离类中心。而当新任务中需要生成 $M_1,\cdots,M_k$ 类示例样本时，只需从对应的 $N(\bar{\boldsymbol{z}}_{y_i},\lambda\sigma_{z_{y_i}})$ 中采样并采用原 VAE 模型解码器重构，即可得到标记样本对 $(\boldsymbol{x}_o',y)$，并与新增类样本对 $(\boldsymbol{x}_n,y)$ 联合训练新模型。

基于知识蒸馏的模块采用软决策向量学习原模型的知识。在模型的输出向量中，即使负标签也带有大量信息。而在以往方法中负标签常常被忽略。为此，本文基于知识蒸馏思想设计了如式(8)所示的损失函数，用于将原模型的知识传递到新模型中。该方法通过拟合 C_{old} 类样本在原模型上的软决策向量来实现。

$$L_{\mathrm{soft}} = -\mathbb{E}\left(\left(\exp\left(\frac{\frac{A(S_o(Q_{\mathrm{E}}^o(\boldsymbol{x}_i)))}{T}}{\sum_j^{C_{\mathrm{all}}}\frac{A(S_o(Q_{\mathrm{E}}^o(\boldsymbol{x}_j)))}{T}}\right)\log\left(\exp\left(\frac{\frac{A(S_n(Q_{\mathrm{E}}^n(\boldsymbol{x}_i)))}{T}}{\sum_j^{C_{\mathrm{all}}}\frac{A(S_n(Q_{\mathrm{E}}^n(\boldsymbol{x}_j)))}{T}}\right)\right)\right) \tag{8}$$

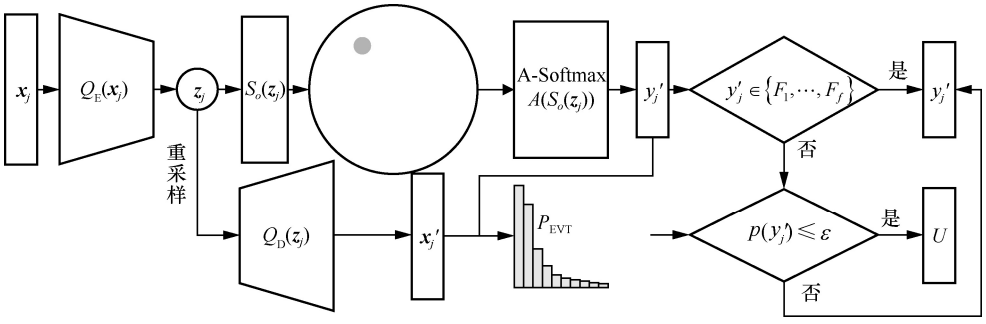


图 2 测试阶段模型的总体流程

其中，Q_{E}^{o}、S_o 为原模型的 VAE 和角投影层函数；Q_{E}^{n}、S_n 为新模型的 VAE 和角投影层函数；T 为温度系数超参数；A 为对应的 A-Softmax 函数。并且对于新模型输出的决策向量 $A\left(S_n(Q_{\mathrm{E}}^{n}(\boldsymbol{x}_i))\right)$，只截取 C_{old} 类部分计算以保持新模型与原模型的输出维度相同。

综上所述，当有新任务时，所需存储的数据仅包括原模型参数、小样本类样本以及原 VAE 模型中存储的 k 个正态分布，存储成本随着新任务数量的增加缓慢增长，满足增量学习的需求。新模型采用大样本恶意类（$M_1,\cdots,M_k$）的重构样本 $\boldsymbol{x}_o'$、新增类样本 $\boldsymbol{x}_n$ 以及良性类 B 和存储的小样本类样本联合训练，其损失函数如式(9)所示，对应于图 1 中的③。

$$L_{\mathrm{All}} = L_{\mathrm{vae}} + \alpha L_{\mathrm{Ang}} + \beta L_{\mathrm{Cont}} + \gamma L_{\mathrm{soft}} \tag{9}$$

其中，α、β 和 γ 都是权重超参数，并且 α、β 可以直接继承原模型式(5)中的参数值。

最后，基于训练后的模型构建 $k+\Delta k+1$ 个 EVT 模型实现对未知类的识别，对应于图 1 中的④。当有更新的分类任务时，该阶段的模型就成为原模型，再次开始迭代学习更新的分类任务。

3　实验分析

本节在一个恶意流量识别文献中广泛使用的公开数据集 NSL-KDD[33]上评估了所提方法。首先，在初始的原任务上将所提方法与现有方法进行对比分析；其次，对所提方法各个组件在增量学习场景下的有效性进行分析验证。实验表明，所提方法在识别大样本类、小样本类和未知类恶意流量识别上均表现出色，并且在识别新增恶意类流量时所提方法能够极大降低对原有类知识的遗忘速度，适用于增量学习场景下的快速识别。

3.1　数据集

实验采用了 NSL-KDD 数据集。该数据集包括良性类和 39 个细粒度恶意类。其中，恶意流量可分为 4 个粗粒度类，即拒绝服务（DoS, denial of service）类、扫描（Probe）类、本地提权（U2R）类和远程（R2L）类，具体如表 1 所示。本文选择 NSL-KDD 主要基于以下两点：1) NSL-KDD 一直被高水平研究文献所采用，便于与其他经典方法相比较；2) 相比其他数据集，NSL-KDD 中包含了更丰富的细粒度类。

表 1　NSL-KDD 数据集

粗粒度类	细粒度类	训练集数量	测试集数量
良性	benign	67 343	9 711
DoS	apache2	0	737
	back	956	359
	land	18	7
	neptune	41 214	4 657
	mailbomb	0	293
	pod	201	41
	processtable	0	685
	smurf	2 646	665
	teardrop	892	12
	udpstorm	0	2
Probe	ipsweep	3 599	141
	mscan	0	996
	nmap	1 493	73
	portsweep	2 931	157
	saint	0	319
	satan	3 633	735
U2R	buffer_overflow	30	20
	httptunnel	0	133
	loadmodule	9	2
	perl	3	2
	ps	0	15
	xterm	0	13
	rootkit	10	13
	sqlattack	0	2
R2L	worm	0	2
	ftp_write	8	3
	guess_passwd	53	1 231
	imap	11	1
	multihop	7	18
	named	0	17
	phf	4	2
	sendmail	0	14
	snmpgetattack	0	178
	snmpguess	0	331
	spy	2	0
	warezclient	890	0
	warezmaster	20	944
	xsnoop	0	4
	xlock	0	9

基于表 1 和实际情况发现，DoS 类和 Probe 类一般存在大量标记样本，U2R 类和 R2L 类则更频繁地存在小样本类或未知类的情况。因此，实验时 DoS 类和 Probe 类大多作为大样本恶意类，U2R 类

和 R2L 类大多作为小样本类和未知类。并且，实验中小样本类中每类仅选取 5 个样本（5-shot）用于训练。

在初始任务上，实验中采用 DoS 类中的 back、neptune、smurf 和 Probe 类中的 ipsweep、satan 以及 R2L 类中的 guess_passwd、warezclient 作为大样本恶意类，U2R 类中的 buffer_overflow、rootkit 以及 R2L 类中的 multihop、warezmaster 作为小样本类。后续出现新任务时，每次随机新增 4 个类（包括 3 个大样本类和一个小样本类）。

3.2 评估标准

分类任务的性能一般采用精度（Precision）、召回率（Recall）和 F_1 值指标衡量。

$$\text{Precision}=\frac{\text{TruePositives}}{\text{TruePositives}+\text{FalsePositives}} \tag{10}$$

$$\text{Recall}=\frac{\text{TruePositives}}{\text{TruePositives}+\text{FalseNegatives}} \tag{11}$$

$$F_1=\frac{2\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

本文同样关注小样本类，因此均采用了宏（Macro）平均指标，宏平均 F_1 值如式(13)所示。

$$F_{1-\text{Macro}}=\frac{1}{\left\|C_{\text{new}}\cup C_{\text{old}}\right\|}\sum_{c_k\in C_{\text{new}}\cup C_{\text{old}}}F_{1-c_k} \tag{13}$$

3.3 结果分析

本节对本文模型中各组件的有效性和模型的先进性进行了分析。首先，分析基模型的分类性能。

在原任务上的已知类识别阶段，本节对比了随机森林和多层感知机（MLP, multilayer perceptron）2 种基线方法、CVAE-EVT[5]、对比 CVAE[12]、所提方法及其变体的检测性能，相关结果如表 2 所示。从表 2 可以看出，所提方法在已知类识别阶段的良性类、大样本已知类上均实现了最佳分类性能，在小样本已知类上也与对比 CVAE[12]接近。并且，本文提出的对比学习以及 A-Softmax 这 2 个模块都对所提方法的细粒度识别性能有较大提升。实验结果证明了所提方法的有效性和先进性。

表 2 原任务上已知类识别阶段不同方法检测性能对比

方法	良性类			大样本已知类			小样本已知类		
	精度	召回率	F_1 值	精度	召回率	F_1 值	精度	召回率	F_1 值
随机森林	0.817	0.954	0.88	0.829	0.780	0.705	0	0	0
MLP	0.758	0.882	0.815	0.533	0.895	0.628	0.376	0.232	0.244
CVAE-EVT	0.871	0.866	0.868	0.561	0.959	0.708	0.348	0.364	0.289
对比 CVAE	0.876	0.903	0.890	0.581	0.960	0.723	0.410	0.567	0.476
所提方法	0.941	0.909	0.925	0.714	0.986	0.793	0.342	0.676	0.410
无对比学习的方法	0.931	0.653	0.768	0.599	0.993	0.702	0.302	0.587	0.329
无 A-Softmax 的方法	0.927	0.871	0.898	0.664	0.956	0.747	0.306	0.445	0.288

原任务上未知类识别阶段所提方法的归一化混淆矩阵如图 3 所示。从图 3 可以看出，所提方法在保持已知类识别阶段分类性能的同时，实现了对未知类的识别。此外，本节对比了 CVAE-EVT[5]、对比 CVAE[12]以及所提方法在原任务上未知类的检测性能（4 种粗粒度未知类上的宏平均指标），识别阶段如图 4 所示。从图 4 可以看出，所提方法对未知类的识别精度虽略逊于基于 CVAE 的方法，但差距并不大。并且由于所提方法是针对增量学习场景设计的，因此未知类可以经由安全人员及时分析后作为新增类以更新模型，使未知类可以快速被转化为已知类，从而实现高精度的细粒度识别。

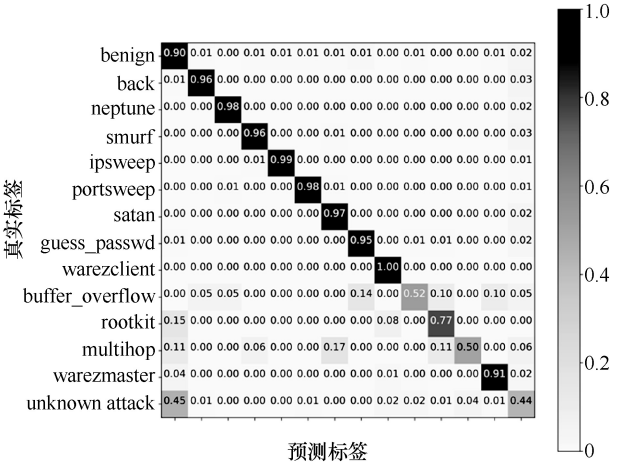


图 3 原任务上未知类识别阶段所提方法的归一化混淆矩阵

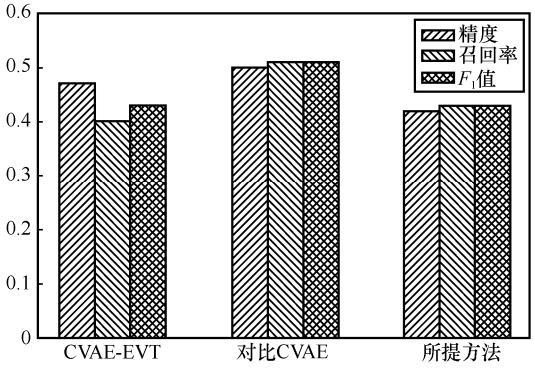
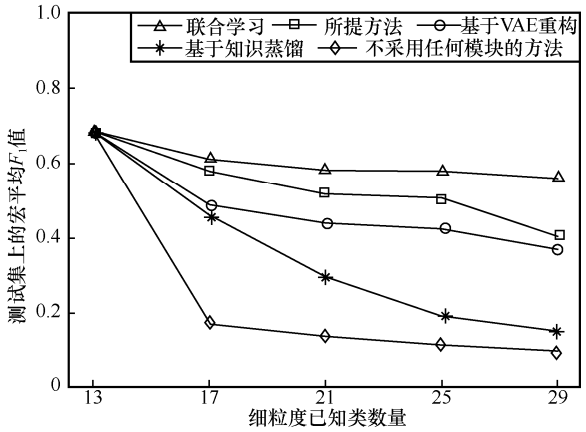
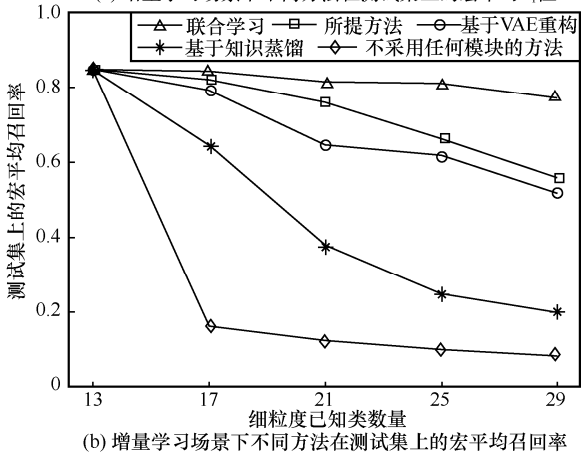


图 4　原任务上未知类识别阶段不同方法检测性能对比

在新增类识别阶段，每次有新任务时模型在上一次原任务的基础上固定训练 50 次，增量学习场景下不同模型在测试集上的细粒度分类性能如图 5 所示。从图 5 可以看出，对于实际数量较少的恶意类而言，召回率更重要。相比联合学习下的性能上限，基于 VAE 重构和基于知识蒸馏的模块都对保留原有类知识有较大提升作用，特别是采用 VAE 重构的模块提升更加明显。实验结果说明，所提方法能够在增量学习场景下实现高性能的细粒度流量识别，这证明了所提方法的有效性。



(a) 增量学习场景下不同方法在测试集上的宏平均F_1值



(b) 增量学习场景下不同方法在测试集上的宏平均召回率

图 5　增量学习场景下不同方法在测试集上的细粒度分类性能

4　结束语

本文旨在设计一种增量学习场景下的细粒度恶意流量识别方法。在对比 CVAE-EVT[12]的基础上，本文基于 VAE 模型提出了一种能兼顾已知类、小样本类和未知类恶意流量的识别方法，并且采用 VAE 重构和知识蒸馏方法，实现了增量学习条件下的新增类识别。所提方法对大样本类、小样本类、未知类和新增类恶意流量都在公开数据集上分类表现优异。该方法在当前未知攻击愈发常见的背景下能迅速实现对新增类的学习，并不需要维持庞大的原有类样本数据库，适用于对检测时间和存储成本有较高需求的应用场景，或作为联合学习模型训练完成之前的过渡模型。最后，由于攻击技术的不断发展，新增类恶意流量或更加接近于良性类流量，从而造成增量学习方法失效，未来拟进一步研究针对此问题的解决方案。

参考文献：

[1] SOYSAL M, SCHMIDT E G. Machine learning algorithms for accurate flow-based network traffic classification: evaluation and comparison[J]. Performance Evaluation, 2010, 67(6): 451-467.

[2] DUSI M, GRINGOLI F, SALGARELLI L. Quantifying the accuracy of the ground truth associated with Internet traffic traces[J]. Computer Networks, 2011, 55(5):1158-1167

[3] 陈明豪, 祝跃飞, 芦斌, 等. 基于 attention-CNN 的加密流量应用类型识别[J]. 计算机科学, 2021, 48(4): 325-332.
CHEN M H, ZHU Y F, LU B, et al. Classification of application type of encrypted traffic based on attention-CNN[J]. Computer Science, 2021, 48(4): 325-332.

[4] TING C, FIELD R, FISHER A, et al. Compression analytics for classification and anomaly detection within network communication[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1366-1376.

[5] YANG J, CHEN X, CHEN S W, et al. Conditional variational auto-encoder and extreme value theory aided two-stage learning approach for intelligent fine-grained known/unknown intrusion detection[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 3538-3553.

[6] CASTRO F M, MARÍN-JIMÉNEZ M J, GUIL N, et al. End-to-end incremental learning[C]//Proceedings of the European Conference on Computer Vision. New York: ACM Press, 2018: 233-248.

[7] LANGE D M, ALJUNDI R, MASANA M, et al. A continual learning survey: defying forgetting in classification tasks[J]. IEEE

Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3366-3385.

[8] BUKCHIN G, SCHWARTZ E, SAENKO K, et al. Fine-grained angular contrastive learning with coarse labels[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8726-8736.

[9] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[10] HADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping[C]//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2006: 1735-1742.

[11] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. arXiv Preprint, arXiv:1503.02531, 2015.

[12] 王一丰, 郭渊博, 陈庆礼, 等. 基于对比学习的细粒度未知恶意流量分类方法[J]. 通信学报, 2022, 43(10):12-25.
WANG Y F, GUO Y B. GHEN Q L, et al. Method based on contrastive learning for fine-grained unknown malicious traffic classification[J]. Journal on Communications, 2022, 43(10):12-25.

[13] KINGMA D P, WELLING M. Auto-encoding variational Bayes[J]. arXiv Preprint, arXiv:1312.6114, 2013.

[14] LIU W Y, WEN Y D, YU Z D, et al. SphereFace: deep hypersphere embedding for face recognition[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6738-6746.

[15] HAAN L D, FERREIRA A. Extreme value theory: an introduction[M]. New York: Springer, 2006.

[16] DONG B, WANG X. Comparison deep learning method to traditional methods using for network intrusion detection[C]//Proceedings of 2016 8th IEEE International Conference on Communication Software and Networks. Piscataway: IEEE Press, 2016: 581-585.

[17] CRUZ S, COLEMAN C, RUDD E M, et al. Open set intrusion recognition for fine-grained attack categorization[C]//Proceedings of 2017 IEEE International Symposium on Technologies for Homeland Security. Piscataway: IEEE Press, 2017: 1-6.

[18] HENRYDOSS J, CRUZ S, RUDD E M, et al. Incremental open set intrusion recognition using extreme value machine[C]//Proceedings of 2017 16th IEEE International Conference on Machine Learning and Applications. Piscataway: IEEE Press, 2018: 1089-1093.

[19] MALLYA A, LAZEBNIK S. PackNet: adding multiple tasks to a single network by iterative pruning[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7765-7773.

[20] LI Z Z, HOIEM D. Learning without forgetting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 2935-2947.

[21] RANNEN A, ALJUNDI R, BLASCHKO M B, et al. Encoder based lifelong learning[C]//Proceedings of IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1329-1337.

[22] ZHANG J T, ZHANG J, GHOSH S, et al. Class-incremental learning via deep model consolidation[C]//Proceedings of IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2020: 1120-1129.

[23] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. ICARL: incremental classifier and representation learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 2001-2010.

[24] ROLNICK D, AHUJA A, SCHWARZ J, et al. Experience replay for continual learning[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2019: 350-360.

[25] SOHN K. Improved deep metric learning with multi-class n-pair loss objective[C]//Proceedings of International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2016: 29.

[26] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of International Conference on Machine Learning. New York: PMLR, 2020: 1597-1607.

[27] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 9726-9735.

[28] HU Q J, WANG X, HU W, et al. AdCo: adversarial contrast for efficient learning of unsupervised representations from self-trained negative adversaries[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 1074-1083.

[29] LASHKARI A H, DRAPERGIL G, MAMUN M S I, et al. Characterization of ToR traffic using time-based features[C]//Proceedings of 3rd International Conference on Information Systems Security and Privacy. New York: ACM Press, 2017: 253-262.

[30] DOERSCH C. Tutorial on variational autoencoders[J]. arXiv Preprint, arXiv:1606.05908, 2016.

[31] HIGGINS I, MATTHEY L, PAL A, et al. Beta-VAE: learning basic visual concepts with a constrained variational frame-work[C]//Proceedings of International Conference on Learning Representations. [S.l.:s.n.], 2017: 1-22.

[32] OORD V D A, LI Y, VINYALS O. Representation learning with

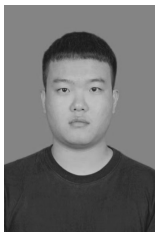
contrastive predictive coding[J]. arXiv Preprint, arXiv: 1807.03748, 2018.

[33] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]//Proceedings of 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. Piscataway: IEEE Press, 2009: 1-6.

[作者简介]

王一丰（1994– ），男，江苏泰兴人，信息工程大学博士生，主要研究方向为零样本学习、网络安全和入侵检测。

郭渊博（1975– ），男，陕西周至人，博士，信息工程大学教授、博士生导师，主要研究方向为大数据安全、态势感知。

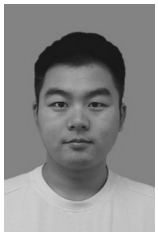


陈庆礼（1998– ），男，河南新乡人，信息工程大学硕士生，主要研究方向为人工智能安全。

方晨（1993– ），男，安徽宿松人，博士，信息工程大学讲师，主要研究方向为机器学习、隐私安全。

林韧昊（1993– ），男，河南郑州人，郑州大学博士生，主要研究方向为深度学习、稳健性验证和网络安全等。

周永良（1983– ），男，河北衡水人，信息工程大学工程师，主要研究方向为网络信息安全、信息化通信技术保障。



马佳利（1996– ），男，福建福清人，信息工程大学博士生，主要研究方向为数字孪生、网络安全。